

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-167110

(43)Date of publication of application : 22.06.2001

(51)Int.Cl.

G06F 17/30

H04N 5/76

(21)Application number : 11-349526

(71)Applicant : MATSUSHITA ELECTRIC IND  
CO LTD

(22)Date of filing : 08.12.1999

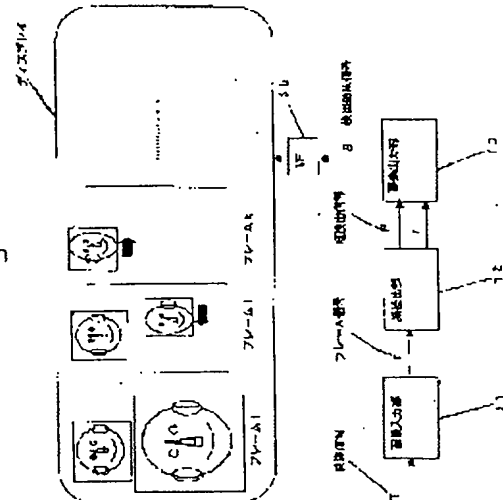
(72)Inventor : ITO RYUTA  
YAMADA SHIN  
SOMA MASAYOSHI  
NAGAO KENJI

## (54) PICTURE RETRIEVING METHOD AND ITS DEVICE

(57)Abstract:

**PROBLEM TO BE SOLVED:** To distinguish and display the faces of persons appearing in video by detecting the faces in particular from the video and identifying the detected face in addition.

**SOLUTION:** The device provided with a means for detecting a face from the video and a means for identifying the detected face detects a frame including the face from the video, extracts a face picture from the frame, and groups the faces of the same appearing person from all the extracted face pictures to extract the representative face picture of each appearing person to identify the face of the appearing person in the video. Thus, the face of the person appearing in the video can be distinguished and displayed.



**\* NOTICES \***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**CLAIMS**

---

[Claim(s)]

[Claim 1]An image retrieval method detecting a frame to which a face is reflected out of an image, carrying out grouping of the face of the same characters from all the face pictures which extracted and extracted a face picture from said frame, and extracting the representation face picture according to characters.

[Claim 2]The image retrieval method according to claim 1, wherein detection of a frame to which a face is reflected from an image detects a break of a scene from an image and detects a frame in which a face is contained as a representative picture image of each scene.

[Claim 3]The number of faces with which detection of a frame to which a face is reflected from an image is reflected (number), The image retrieval method according to claim 1 or 2 detecting at least one or more of size of a face, direction of a face, sex, expression of a face, existence of glasses (age presumption, racial judging of a face), or mustached existence as predetermined conditions.

[Claim 4]An image retrieval method detecting a frame which calculated similarity with a face picture specified by a face and a retrieving person of characters in an image, and to which a face more than predetermined similarity was reflected.

[Claim 5]The image retrieval method according to claim 4, wherein a face of characters in an image chooses a frame which detects a break of a scene from an image and in which a face is contained as a representative picture image of each scene.

[Claim 6]An image retrieval method extracting a frame to which a posture of a person who detects from an image a frame to which a person is reflected, and fulfills predetermined conditions, or a dress was reflected.

[Claim 7]The image retrieval method according to claim 6, wherein detection of a frame to which a person is reflected from an image detects a break of a scene from an image and detects a frame in which a face is contained as a representative picture image of each scene.

[Claim 8]The image retrieval method according to claim 4, wherein a face picture specified by a retrieving person specifies at least one or more face pictures from a face database beforehand registered with a characters list.

[Claim 9]The image retrieval method according to claim 8 a retrieving person's creating a characters list beforehand, or generating it from a race card.

[Claim 10]The image retrieval method according to claim 4, wherein a face picture specified by a retrieving person carries out specification registration of the face picture of characters in an image beforehand.

[Claim 11]An image retrieval apparatus comprising:

A face primary detecting element which detects a frame to which a face is reflected out of an image, and extracts at least one or more face pictures from said frame.

A characters identification part which carries out grouping of the face of the same characters from said all extracted face pictures, and extracts the representation face picture according to characters.

[Claim 12]An image retrieval apparatus comprising:

A scene change primary detecting element which detects an end change of a scene by supervising a frame image to a time series by considering a video signal as an input.

A face primary detecting element which detects a frame to which a face is reflected from a representative picture image of each scene from said scene change primary detecting element, and extracts at least one or more face pictures from said frame.

A characters identification part which carries out grouping of the face of the same characters from said all extracted face pictures, and extracts the representation face picture according to characters.

[Claim 13]An image retrieval apparatus comprising:

A characters specification part which specifies at least one or more face pictures from a face database which a retrieving person registered beforehand with a characters list.

A video input section outputted per frame in response to the fact that [ video signal ] as an input.

A face primary detecting element which detects a face area from a frame image outputted from a video input section.

A characters identification part which judges whether it corresponds to characters from a face picture outputted from said characters specification part, and a face picture detected from said face primary detecting element in a face area, and an image output part which displays and records a discriminated result judged from said appearance identification part.

[Claim 14]An image retrieval apparatus comprising:

Face image read in part which reads a picture of a retrieval object.

A video input section which outputs a video signal per frame in response to the fact that [ video signal ] as an input.

A face primary detecting element which detects a face area from a frame image outputted from a video input section.

A collating part which judges whether said retrieval object image and a face picture detected in a face area outputted from said face primary detecting element are retrieval object \*\*\*\*, and an image output part which displays and memorizes said collated result.

---

[Translation done.]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.\*\*\*\* shows the word which can not be translated.

3.In the drawings, any words are not translated.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention]This invention relates to the method and device which search for the picture to which the specific goods in video are reflected in a video editing device or an image retrieval apparatus.

[0002]

[Description of the Prior Art]What was indicated to JP,6-223179,A is known as a conventional image retrieval apparatus. The above-mentioned application is an image retrieval apparatus which detects the frame to which specific goods (retrieval object) are reflected from color video. Drawing 1 is a block diagram showing the composition. The computer as a central processing unit with which 1 judges a specific subject in drawing 1. The display of CRT etc. as which 2 displays the output picture of the computer 1. The A/D converter which changes 3 into video playback equipment, such as an optical disc, and from which 4 changes an analog signal into a digital signal. The control line with which 5 connects the control signal between the computers 1 with the video playback equipment 3, the external storage with which 6 consists of hard disks etc., and 7 input devices, such as a mouse, and 8a-8e. The interface which makes connection between the computer 1 and peripheral equipment, CPU which performs data processing [ in / in 9 / a computer ], and 10 are memories which carry out direct access from CPU9.

[0003]Drawing 12 is a flow chart which shows operation of the above-mentioned conventional image retrieval apparatus. Hereafter, operation is explained according to the flow chart of drawing 12. First, if the picture which a retrieving person wants to search is chosen and it inputs into a device from the input device 7 etc., one picture including a subject will be specified (Step 2001). This device is divided into the field of a similar color about the inputted picture (Step 2002). The histogram of a color is generated about each divided subregion (Step 2004), N colors with high frequency are chosen in order, and list CG (r) of Table 1 is created (Step 2005). However, CG (r) shows the list of the r-th subregion. List CG of the subregion which furthermore adjoins list CG (r) and the subregion concerned by the physical relationship in a picture about each subregion The list RCGP which shows a correspondence relation with (r) in Table 2 is created (Step 2007).

[0004]Next, operation of search is explained. Drawing 13 is a flow chart which shows a search operation. First, plurality does cell C (x, y) division of the frame of a determination object (201), a color histogram is generated about each cell (Step 203), and a color with larger frequency of a histogram than a set threshold is registered into list CC (x, y) (Steps 205 and 206). next, color group CG which constitutes \*\*\*\* in (Step 208) and RCGP about each cell C (x, y) — one belonging to one inner color group of colors, One which is contained in list CC (x, y) (Step 209), and belongs to the color group of another side of colors, When contained in list CC of the cell of either the cell C (x, y) itself or its eight adjoining cells, the cell C (x, y) is extracted as an effective cell (Step 210), and it asks for the total of the effective cell in one frame (Step 211). If the number of the effective cells in one frame is beyond a threshold (Step 212) A color group versus the CG is counted as an effective color group pair (Step 213). And it will be a subject if the counted total of

an effective color group pair becomes in more than a set threshold. (retrieval object) It judges with existing in a frame (Step 215). If below a threshold becomes, it will judge as that in which a subject does not exist (Step 216).

[0005]In this image retrieval apparatus, the burden of a video editor or an image retrieval person is made light, and the art which discovers the picture of a retrieval object efficiently is demanded. However, since it refers to the above-mentioned conventional art based on the information on the physical relationship between color-group-izing and the color group of a picture to search, For example, when dresses, such as a uniform, are the same, it has the technical problem that it will not judge with his being a person, and that the increase in erroneous detection will be invited to a retrieval object, irrespective of that of different person \*\*\*\* from the person who wants to search in practice.

[0006]In this invention, it was made in view of the above-mentioned conventional problem. Therefore, an aforementioned problem is solved by identifying the face which performed detection which specialized in the face and was further detected from the image.

[0007]

[Means for Solving the Problem]What makes it a gist to have enabled it to identify that this invention is a person characterized by comprising the following who is different more by being alike even if there is little erroneous detection and it is the same dress in order to solve this technical problem.

A means to detect a face from an image.

A means to identify a detected face.

[0008]As an invention which has such a mode, an invention of a statement to this invention claim 1, As an image retrieval method, a frame to which a face is reflected is detected out of an image, Extract a face picture from said frame and grouping of the face of the same characters is carried out from all the extracted face pictures, The representation face picture is extracted according to characters, and it is considered as an image retrieval apparatus identifying a face of characters in an image, and has the operation that a face of a person who appears into an image can be distinguished and displayed.

[0009]Detection of a frame to which a face is reflected from an image in the image retrieval method according to claim 1 as for the invention of this invention according to claim 2, A break of a scene is detected from an image, a frame in which a face is contained as a representative picture image of each scene is detected, and it has the operation that a picture to which a face is reflected as a representative picture image of each scene can be chosen.

[0010]Detection of a frame to which a face is reflected from an image in the image retrieval method according to claim 1 or 2 as for the invention of this invention according to claim 3, The number of reflected faces (number), size of a face, direction of a face, sex, expression of a face, (Age presumption, racial judging of a face) At least one or more of existence of glasses or mustached existence are detected as predetermined conditions, It has the operation that a picture in which a face picture which is in agreement with conditions which a retrieving person specified was included can be extracted, by extracting a face picture by which predetermined conditions are fulfilled.

[0011]The invention of this invention according to claim 4 calculates similarity with a face picture specified by a face and a retrieving person of characters in an image as an image retrieval method, A frame to which a face more than predetermined similarity was reflected is detected, and it has the operation that it can be known where [ in an image ] a face specified by a retrieving person is recorded.

[0012]In the image retrieval method according to claim 4 the invention of this invention according to claim 5, A break of a scene is detected from an image, and it is considered as an image retrieval apparatus choosing a frame in which a face is contained as a representative picture image of each scene, and has the operation that a picture to which a face is reflected as a representative picture

image of each scene can be chosen.

[0013]The invention of this invention according to claim 6 detects from an image a frame to which a person is reflected as an image retrieval method, It has the operation that a frame to which a posture of a person who fulfills predetermined conditions, or a dress was reflected is extracted, and a picture in which a specific person's attribute was contained can be extracted.

[0014]Detection of a frame to which a person is reflected from an image in the image retrieval method according to claim 6 as for the invention of this invention according to claim 7, A break of a scene is detected from an image, a frame in which a face is contained as a representative picture image of each scene is detected, and it has the operation that a picture to which a face is reflected as a representative picture image of each scene can be chosen.

[0015]A face picture as which a retrieving person specifies the invention of this invention according to claim 8 in the image retrieval method according to claim 4 specifies at least one or more face pictures from a face database beforehand registered with a characters list, and has the operation that a person can be specified easily.

[0016]In the image retrieval method according to claim 8, the invention of this invention according to claim 9 a characters list, A retrieving person creates beforehand or it is made to generate from a race card, and while being able to perform specification of a face picture by a retrieving person simply by creating beforehand, it has the operation that a characters list can generate simply and easily, by generating from a race card.

[0017]A face picture as which a retrieving person specifies the invention of this invention according to claim 10 in the image retrieval method according to claim 4, It is made to carry out specification registration of the face picture of characters in an image beforehand, and has the operation that specification of a face picture by a retrieving person can be performed simply, by carrying out specification registration of the face picture beforehand.

[0018]A face primary detecting element which the invention of this invention according to claim 11 detects a frame to which a face is reflected out of an image as an image retrieval apparatus, and extracts at least one or more face pictures from said frame, Grouping of the face of the same characters is carried out from said all extracted face pictures, and it has a characters identification part which extracts the representation face picture according to characters, and has the operation that a face of a person who appears into an image can be distinguished and displayed.

[0019]A scene change primary detecting element which detects an end change of a scene by the invention of this invention according to claim 12 considering a video signal as an input as an image retrieval apparatus, and supervising a frame image to a time series, A face primary detecting element which detects a frame to which a face is reflected from a representative picture image of each scene from said scene change primary detecting element, and extracts at least one or more face pictures from said frame, Grouping of the face of the same characters is carried out from said all extracted face pictures, and it has a characters identification part which extracts the representation face picture according to characters, and has the operation that it can display on a display as a frame representative picture image in which a face is reflected for every scene, or can memorize.

[0020]A characters specification part with which the invention of this invention according to claim 13 specifies at least one or more face pictures as an image retrieval apparatus from a face database which a retrieving person registered beforehand with a characters list, A video input section outputted per frame in response to the fact that [ video signal ] as an input, A face primary detecting element which detects a face area from a frame image outputted from a video input section, A characters identification part which judges whether it corresponds to characters from a face picture outputted from said characters specification part, and a face picture detected from said face primary detecting element in a face area, It has an image output part which displays and records a discriminated result judged from said appearance identification part, and has the operation that characters can be displayed on a display according to characters, or can be memorized to a certain video program.

[0021]Face image read in part into which the invention of this invention according to claim 14 reads a picture of a retrieval object as an image retrieval apparatus, A video input section which outputs a video signal per frame in response to the fact that [ video signal ] as an input, A face primary detecting element which detects a face area from a frame image outputted from a video input section, A collating part which judges whether said retrieval object image and a face picture detected in a face area outputted from said face primary detecting element are retrieval object \*\*\*\*, a face picture which was provided with an image output part which displays and memorizes said collated result, and a retrieving person specified -- coincidence -- better -- \*\* has the operation that a frame including a similar face picture can be displayed on a display, or can be memorized.

[0022]

[Embodiment of the Invention](Embodiment 1) An embodiment of the invention is hereafter described using drawing 11 from drawing 1. Drawing 1 thru/or drawing 3 are the figures explaining the image retrieval apparatus concerning a 1st embodiment of this invention. This embodiment explains detection of a face. When it judges whether a face specifically exists in the read picture and a face area exists, direction (a transverse plane, facing up, facing down, facing the left, facing the right) of the position of the face in a picture, the number and the size of a face, and a face distinguishes, and the result is displayed on a monitor etc.

[0023]Drawing 1 is an entire configuration figure of the sensing device of a face. In drawing 1, 1 is a computer which detects an object image, and 2 is CRT etc. which display the detection result of the computer 1, 3 is video playback equipment, such as an optical disc, and 4 is an A/D converter which changes an analog signal into a digital signal, The control line with which 5 connects the control signal between the computers 1 with the video playback equipment 3, 6 is an external storage which consists of hard disks etc., and the interface whose 8a-8e 7 performs input devices, such as a mouse and a keyboard, and make connection between the computer 1 and peripheral equipment, and 9 are CPU of a computer, and a memory which carries out direct access of ten from CPU9.

[0024]One by one, the video signal outputted from the video playback equipment 3 is changed into a digital image by A/D converter 4, and is sent to the computer 1 by it. By computer 1, a digital image goes into the memory 10 via the interface 8c, and is processed by CPU9 according to the program stored in the memory 10.

[0025]Drawing 2 is the flow chart which showed the flow of the detecting method. First, if frame image r is read into the computer 1 (Step 101), a face area will be detected by the function beforehand described by the program according to f (Step 102). The output P of the function f is the procession on which all the face coordinates detected on the frame were recorded. For example, with the frame r, store the vector of eye i sequence of the procession P, and the face detected by the i-th the ingredient (i, 1) of the procession P, For example, if a face area is started in a rectangular field, as for the ingredient (i, 2), the x-coordinate at the upper left of a face area stores the y-coordinate at the upper left of a face area (i, 3), and an ingredient, and the x-coordinate at the lower right of a face area and the ingredient (i, 4) store the y-coordinate at the lower right of a face area. When a face area does not find one, the function f returns -1. If the output of the procession P is not -1, a detection result will be outputted to the display 2 (Step 104).

[0026]Only the detected face area may be displayed on a display, a specific mark may be given to a face area and the whole frame may be displayed on it. Since the number which the sequence size of the procession P detected is expressed, the detection number can also be collectively displayed on a display.

[0027]In the case of a color picture, about the concrete function f, it is realizable by beige detection, for example. The value (R, G, B) of each pixel of the inputted picture is plotted to a color space (for example, Yuv space), and only the pixel contained in the beige space defined beforehand is chosen. Face detection is realized the same as that of the shape (for example, ellipse) which is a field with the area beyond the threshold beforehand set up among the imaging

ranges formed only by the above-mentioned applicable pixel, and the shape of the field defined beforehand, or by choosing only a similar field. The method by template matching is also considered as a method of responding to both a monochrome picture and a color picture. One sheet or two or more face pictures are beforehand memorized as a standard pattern. To the frame image, some fields on a frame are started one by one, moving the position of the window for logging, and the similarity of a logging picture and an above-mentioned template is calculated. For example, when the correlation value of a logging picture and a template picture is more than a set threshold, it judges with the logging field being a face picture.

[0028]It is possible by preparing the template of different size to also detect the face area where the sizes of a face differ. The face area where direction of a face differs is also detectable by preparing the template from which direction of a face differs. For example, if it is beyond the threshold which prepares the template of facing up and four kinds of downward, leftward, and rightward face pictures, and asked for similarity with the four above-mentioned kinds of templates from each logging picture, and the maximum of these four similarity defined beforehand, Direction of a face can be specified from an applicable template.

[0029]As a realization method of another function f, the way the started field creates beforehand that discriminant function which is a face or a non-face is also considered. For example, by study, the discriminant function of a face and a non-face is created using a neural network. A neural network can be made to be able to learn so that the category formed by beforehand different waving to the picture from which direction of a face differs may be identified, and direction of a face can also be detected.

[0030]Drawing 3 shows processing of face detection with a block diagram. In drawing 3, 11 is an image input part, 12 is a face primary detecting element, and 13 is an image output part. An image input part outputs the picture for one frame as frame signal r by considering the video signal T outputted from A/D converter 4 as an input. The face primary detecting element 12 outputs the face detecting signal P and frame signal r which output the coordinates of the position of the face which exists in one frame, and the information on direction of a face by considering frame signal r outputted from the image input part as an input. The image output part 13 outputs the picture of only a face area as the detection result signal S from the face detecting signal P and frame signal r which are outputted from the face primary detecting element 12. The detection result signal S is displayed on the display 2 via the interface 8b in a computer. The face picture started, for example for every frame is displayed on the frame of drawing 2. Like drawing 2, it can respond to the size of various faces, and direction. For example, direction of a face may be expressed as an arrow etc. on a display. The picture of the whole frame to which the specific mark was given by the face area may be sufficient as the detecting signal S. A mustache, detection of glasses, etc. can be performed with a technique like face detection.

[0031](Embodiment 2) Drawing 4 is a flow chart explaining operation of the image retrieval apparatus concerning a 2nd embodiment of this invention. The image retrieval apparatus concerning this embodiment has the same composition as the equipment configuration concerning a 1st embodiment of the above. According to a 2nd embodiment, when you choose the representative picture image in a certain scene combining the face detection function described by the scene change function and Embodiment 1 of the image, let the picture to which people's face is reflected be a representative picture image. The flow chart shown in drawing 4 below explains.

[0032]In drawing 4, the flag Flg of whether the representative picture image of the present scene is already memorized first is initialized (Step 107). Next, the r-th frame r of an image is read and it is judged whether the scene changed or not (Step 109). representative picture image RT of the front scene beforehand memorized when it is judged that the scene changed -- a display -- displaying (Step 110) -- the flag Flg is reset to zero (Step 111). However, immediately after [ no ] the start of a video input, since representative picture image RT of a front scene does not exist, it is displayed on a display.

[0033]Next, the flag Flg is evaluated (Step 112), when Flg is 0, the frame r read now is recorded as



representative picture image RT (Step 113), and the flag Flg is set to one (Step 114). This processing is a thing in consideration of the case where people's face is not once contained in the applicable scene, and memorizes the frame immediately after judging a scene change as a representative picture image beforehand. After a scene change is judged, the frame after prescribed frame progress may be beforehand memorized as a representative picture image.

[0034] Since the former frame will already have recorded the representative picture image when the flag is set to 1, face detection is performed as it is. Even after the first representative picture image is incorporated at Steps 113 and 114, \*\*\*\*\* of a face is performed by the technique described to Embodiment 1 (Step 115).

[0035] The following frame is read when it is the output P-1 of the function f as a result of detection (i.e., when a face is not detected) (Step 116). When P is not -1 (i.e., when a face is detected by the frame r read now), a representative picture image is updated (Step 117).

[0036] When the face area exists in a certain scene with the above procedure, the frame to which the face is reflected can be used as a representative picture image. Since the technique described by Embodiment 1 can also judge direction of a face, size, and a number, it can use the frame of a transverse-plane face as a representative picture image, for example, can use a frame with larger size of a face as a representative picture image, or can also choose a frame with more faces as a representative picture image.

[0037] Drawing 5 expresses the above-mentioned procedure with a block diagram. In drawing 5, 11 is an image input part and the same as what was described by Embodiment 1. 14 is a scene change primary detecting element, and 12 is a face primary detecting element and is the same as what was described by Embodiment 1. 13 is an image output part and the same as what was described by Embodiment 1.

[0038] The image input part 11 outputs the picture for one frame as frame signal r by considering the video signal T outputted from A/D converter 4 as an input. In the scene change primary detecting element 14, an end change of a scene is judged from the discontinuity of an image, and scene switching signal C and frame signal r are outputted. The face primary detecting element 12 outputs the face detecting signal P and frame signal r which output the coordinates of the position of the face which exists in one frame, and the information on direction of a face by considering as an input scene switching signal C and frame signal r which were outputted from the scene change primary detecting element 14. The image output part 13 outputs the frame to which the face is reflected as the detection result signal S from the face detecting signal P and frame signal r which are outputted from the face primary detecting element 12. The detection result signal S is displayed on the display 2 via the interface 8b in a computer. When the representative picture image for every scene is displayed on the display 2, for example and the face picture is included in the scene, the picture in which the face is contained can be used as a representative picture image.

[0039] realization of a scene change primary detecting element -- literature "video-index creation edit art", (Matsushita Technical Journal Vol.44 No.5), and Yamada growth others -- it is realizable by using the indicated known art.

[0040] (Embodiment 3) Drawing 6 is a flow chart explaining operation of the image retrieval apparatus concerning a 3rd embodiment of this invention. The image retrieval apparatus concerning this embodiment has the same composition as the equipment configuration concerning a 1st embodiment of the above. This embodiment detects a face about the read frame, and explains implementation of discernment of sex from the detected face.

[0041] It explains using the flow chart of drawing 6. First, if frame image r is read into the computer 1 (Step 119), a face area will be detected by the function beforehand described by the program according to f (Step 120). The function f is the same as what was explained by Embodiment 1. When it is not the output P-1 of the function f (i.e., when a face area is detected), the face of a detection area performs discernment of a male and a woman (Step 122). The function g is a man-and-woman discriminant function. The function g carries out man-and-woman discernment by making the coordinates P and the frame r of a face area on a frame into an argument, and

stores a discriminated result in the output Q. The output Q serves as a procession and the vector of eye i sequence of the procession Q is recorded about the face detected to the i-th. the x-coordinate at the upper left of the face area which the ingredient (i, 1) of the procession Q detected to the i-th, and an ingredient (i, 2) -- a y-coordinate and an ingredient (i, 3) -- as for an ingredient (i, 4), in the y-coordinate at the lower right of a face, in the case of a male, the x-coordinate at the lower right of a face is stored, and, as for an ingredient (i, 5), in the case of +1 and a woman, -1 is stored.

[0042]As a realization method of the function g, sex can collect many known face pictures beforehand, and it can realize with statistical methods, such as discriminant analysis, based on these collected face pictures. For example, a discriminant function is also realizable by a neural network's study. Finally, a detection result is outputted to the display 2 (Step 123).

[0043]Therefore, in this embodiment, after carrying out face detection, the face detected further can display a male or a woman. If it has inputted via [ input device 7 ] the purport that a video editor and an image retrieval person want to search only specific sex beforehand, the frame in which the face picture only according to the specificity was included is also detectable.

[0044]Drawing 7 expresses the above-mentioned processing as a block diagram. In drawing 7, 11 is an image input part, 12 is a face primary detecting element, 18 is a man-and-woman identification part, and 13 is an image output part. The image input part 11 outputs the picture for one frame as frame signal r by considering the video signal T outputted from A/D converter 4 as an input. The face primary detecting element 12 outputs the face detecting signal P and frame signal r which output information, including the coordinates of the position of the face which exists in one frame, direction of a face, size, etc., by considering frame signal r outputted from the image input part as an input. In a man-and-woman identification part, the face detecting signal P and frame signal r which were outputted from the face primary detecting element 12 are considered as an input, man-and-woman discernment is performed based on the picture of the face area recorded on P, a discriminated result and a result with the face detecting signal P are combined, and it outputs as the man-and-woman recognition signal Q. Frame signal r is also outputted. From the man-and-woman recognition signal Q and frame signal r which were outputted from the man-and-woman identification part 15, the image output part 13 outputs the picture which added the identification marking of the male and the woman to each face picture as the detection result signal S, after starting only a face area. The detection result signal S is displayed on the display 2 via the interface 8b in a computer. The sex mark of a face picture and each face picture started, for example for every frame is attached and displayed on the display of drawing 7.

[0045]The way of a display on a display is not what was restricted to what was mentioned above, and after a specific mark is attached by the face area, the whole frame may be displayed on a display. The representative picture image of each scene can be made the face to which characteristic sex is reflected by inserting the scene change primary detecting element 14 which described to Embodiment 2 immediately after the image input part 11 of drawing 7.

[0046]As shown in the block diagram of drawing 8, finer search and edit are attained by adding the age identification part 16, the expression identification part 17, and the racial identification part 18. From the face detecting signal P outputted from the face primary detecting element 12, and frame signal r, the age identification part 16 presumes age from the detected face, and outputs the age recognition signal y. Discernment of age outputs ages, such as his twenties, for example. As a realization method of the age discernment 16, it is realizable in the technique of realizing a man-and-woman identification part, and a similar way.

[0047]An age can collect many known face pictures beforehand, and it can realize with statistical methods, such as discriminant analysis, based on these collected face pictures. For example, a discriminant function is also realizable by a neural network's study. This expression identification part 17 presumes expression from the detected face by considering as an input the face detecting signal P and frame signal r which were outputted from the face primary detecting element 12, and outputs the expression recognition signal H. The label in which discernment of expression was attached according to expression, such as "it laughs", "it crying", and "getting angry", for

example is outputted. As a realization method of the expression identification part 17, it is realizable in the technique of realizing a man-and-woman identification part, and a similar way.

[0048]Expression can collect many known face pictures beforehand, and it can realize with statistical methods, such as discriminant analysis, based on these collected face pictures. For example, a discriminant function is also realizable by a neural network's study.

[0049]The racial identification part 18 presumes a race from the detected face by considering as an input the face detecting signal P and frame signal r which were outputted from the face primary detecting element 12, and outputs the racial recognition signal L. The label in which a race's discernment was attached according to races, such as "yellow-skinned races", a "white", and a "black person", for example is outputted. As a realization method of the expression identification part 17, it is realizable in the technique of realizing a man-and-woman identification part, and a similar way.

[0050]A race can collect many known face pictures beforehand, and it can realize with statistical methods, such as discriminant analysis, based on these collected face pictures. For example, a discriminant function is also realizable by a neural network's study.

[0051]The expression recognition signal H with which the image output part 13 was outputted from the age recognition signal y outputted from the man-and-woman recognition signal Q and frame signal r which were outputted from the man-and-woman identification part 15, and an age identification part, and the expression identification part 17. In response to the fact that [ the racial recognition signal L outputted from the racial identification part ] as an input, the label of sex, age, expression, and a race is given to each face picture, and it is made to display on the display 2 via interface I/F 8b in a computer.

[0052]In addition, by the same technique, a hairstyle, discernment of a hat, etc. are possible. The information on others about a person which were detected is also acquirable by starting the picture around the started face picture. For example, the color of the clothes which the person wears, the color of a necktie, the kind of clothes, etc. are discriminable from the picture located under a face picture. Furthermore, the person's posture, etc. can be detected from techniques, such as background difference, and operation can also be presumed by catching the person's posture change from two or more continuous frames.

[0053]what the scene change primary detecting element 14 which described a 2nd embodiment is inserted for immediately after the image input part 11 of drawing 8 -- the representative picture image of each scene -- according to specificity -- and a specific age -- and specific expression -- and it is possible to make it the face to which the specific person kind is reflected etc.

[0054]Although image edit and a retrieval device were assumed as a use of the embodiment described here, it can use for the buyer analysis not only in the use which nothing restricted to it but a retail store, etc. For example, the information on the relation between a purchasing commodity, and sex and age is acquirable by installing a camera in the register of retail stores, such as a supermarket, detecting a face from the image of the buyer who ranked with the register, and carrying out man-and-woman discernment, age discernment, etc. It is also possible art to contribute to a sales improvement by changing stock on hand using these information.

[0055](Embodiment 4) Drawing 9 is a flow chart explaining operation of the image retrieval apparatus concerning a 4th embodiment of this invention. Also in this embodiment, an image retrieval apparatus has the same composition as the equipment configuration concerning a 1st embodiment of the above. This embodiment explains discernment of the characters within a certain program. The function which the characters discernment said here distinguishes the person who appears in a certain program, and displays is said.

[0056]First, the list of persons who appear in a certain video program is known, and the following is explained supposing the case where the face picture of the characters can take out from an image database etc. Operation is explained according to the flow chart of the processing shown in drawing 9. First, a program characters list is read (Step 126). As for this list, the ID number is given according to each characters, and the stored address of this ID number and a face image database is recorded. Next, a face picture is read from the face database registered for every ID

number. A face picture assumes that the picture from which direction, size, and expression differ, for example is registered.

[0057]Characteristic quantity is extracted from two or more face pictures according to each characters as mentioned above (Step 128). For example, as the realization technique of a characteristic quantity extraction method, KL development occurs, for example. That is, KL development is carried out for every face picture according to characters.

[0058]Even this is preparation for characters discernment.

[0059]Next, the target program image is read (Step 129) and a face picture is detected with the technique described to Embodiment 1 for every frame (Step 130). About the detected face picture, face identification is carried out and matching with characters is performed (Step 131). A subspace method etc. can be used as the face identification technique. Finally, a characters discriminated result is displayed on a display (Step 132).

[0060]Drawing 10 shows the block diagram of the above-mentioned processing. In drawing 10, 19 is a characters list, and 20 is a face database with which the face picture of characters is recorded, 21 is a face image taking part and 22 is a feature extraction part which extracts the feature from a face picture, 23 is a face identification part which identifies the face of characters, and 11 receives a video signal as an input. It is an image input part which outputs frame signal  $r$  for every frame, 12 is a face primary detecting element, 13 is an image output part which outputs a discriminated result, 8b is an interface in a computer, and 2 is a display which displays a discriminated result. Hereafter, operation is explained according to each block.

[0061]the image taking part 21 captures the face image of the characters memorized by the face database from the characters list by which the measure input was carried out, and outputs an applicable face picture as the face database signal  $ft$ . The feature extraction part 22 extracts characteristic quantity from the face database signal  $ft$  outputted from a face image taking part. For example, the characteristic vector which asked by performing KL development according to each characters is outputted as the characteristic quantity signal  $K$ .

[0062]On the other hand, the image input part 11 incorporates the video signal  $T$  inputted via A/D converter 4, and outputs it one frame at a time as frame signal  $r$ . Operation of the face primary detecting element 12 is the same as what was indicated to Embodiment 1. The face area in a frame is detected from frame signal  $r$  outputted from the image input part 11 with the technique described to Embodiment 1, and the face detecting signal  $P$  and frame signal  $r$  are outputted. The characters identification part 23 receives the above-mentioned characteristic quantity signal  $K$ , the face detecting signal  $P$ , and frame signal  $r$  as an input, and starts and vectorizes a face area from the face detecting signal  $P$  and frame signal  $r$  first. As opposed to this vector, a subspace method etc. are carried out, and the started face picture asks for whether it is most similar to which characters, combines the position on the ID number of those characters with the highest similarity, and the frame of an applicable face picture, and outputs as the identified result signal  $Res$ . Frame signal  $r$  is also outputted simultaneously.

[0063]From the identified result signal  $Res$  and frame signal  $r$  which were outputted from the characters identification part 23, the image output part 13 is outputted as the identified result signal  $S$  so that a face picture may be displayed according to characters on the display 2. The method of a display on a display cannot be restricted to this, the whole frame can be displayed on a display, and a different mark according to characters can also be attached and displayed.

[0064]The detection result of characters can also be outputted for every scene by inserting the scene change primary detecting element 14 which described to Embodiment 2 immediately after the image input part 11.

[0065]Although this example explained the case where the list of persons who appear in a video program beforehand was known, when characters are not known, after carrying out face detection with the whole target image first, it is realizable by carrying out teacher-less clustering. To detect where [ in a subsequent image ] the face of the characters in an early scene has appeared by the case where characters are not registered into a face database. A retrieving person and an editor start characters out of the initial scene of an image, and it can realize by inputting into the feature

extraction part 22.

[0066](Embodiment 5) Drawing 11 is a block diagram showing the composition of the image retrieval apparatus concerning a 5th embodiment of this invention. This embodiment explains where [ in an object image ] a certain specific person's face exists, and detecting.

[0067]In drawing 11, 25 is a face picture of a retrieval object, 21 is a face image taking part which captures the face image of a retrieval object, and 22 is a feature extraction part which extracts the feature from a face picture, and 11, Are an image input part, and 12 is a face primary detecting element, and 26, It is a feature extraction part for video signals which extracts characteristic quantity from the face picture detected from the face primary detecting element 12, 24 is a collating part, 13 is an image output part which outputs a discriminated result, 8b is an interface in a computer, and 2 is a display which displays a discriminated result.

[0068]Hereafter, the operation is explained for every block. The image taking part 21 captures the face image 25 of a retrieval object, and performs pretreatment. For example, histogram smoothing etc. are carried out and it outputs as the pretreated retrieval picture signal TTa. The feature extraction part 22 extracts characteristic quantity for predetermined characteristic quantity from the pretreated retrieval picture signal TTa outputted from the face image taking part 21.

[0069]On the other hand, the image input part 11 incorporates the video signal T inputted via A/D converter 4, and outputs it one frame at a time as frame signal r. In the feature quantity extracting part 26 for video signals from frame signal r outputted from the image input part 11, operation of the face primary detecting element 12 detects the face area in a frame, outputs the face detecting signal P and frame signal r, and the outputted signal, After pretreating by starting a face picture, the characteristic quantity signal k and frame signal r are outputted. In the collating part 24, the characteristic quantity signal outputted from the feature extraction part 22 and the feature quantity extracting part for video signals is compared, if it is more than predetermined similarity, it will judge with it being a face of a retrieval object, and the frame will be outputted to the image output part 13. The image output part 13 outputs the inputted frame with a frame number on the display 2.

[0070]According to an above embodiment, a famous actor, a politician, etc. can be searched quickly, for example, and \*\* which makes the burden of a video editor or a retrieving person light is made.

[0071]

[Effect of the Invention]Since a face picture is detected and the detected face is discriminated from the video data base etc. which are accumulated in large quantities according to this invention as mentioned above, It has the advantageous effect that the picture in which the face picture to search with sufficient accuracy was included can be extracted without carrying out erroneous detection even if a uniform etc. are when a dress is the same. It has the effect that it can cater also to a retrieving person's finer retrieval required, such as man-and-woman discernment, as an effect of this invention.

---

[Translation done.]

**\* NOTICES \***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**DESCRIPTION OF DRAWINGS**

---

[Brief Description of the Drawings]

[Drawing 1] The block diagram showing the entire configuration of the image retrieval apparatus concerning a 1st embodiment of this invention

[Drawing 2] The flow chart for explaining operation of Embodiment 1

[Drawing 3] The block diagram for describing said Embodiment 1

[Drawing 4] The flow chart for explaining operation of the image retrieval apparatus concerning a 1st embodiment of this invention

[Drawing 5] It is a block diagram of \*\* in order to describe said Embodiment 2.

[Drawing 6] The flow chart for explaining operation of the image retrieval apparatus concerning a 3rd embodiment of this invention

[Drawing 7] The block diagram for describing the embodiment of the invention 3

[Drawing 8] The block diagram for describing said Embodiment 3

[Drawing 9] The flow chart for explaining operation of the image retrieval apparatus concerning a 4th embodiment of this invention

[Drawing 10] The block diagram for describing said Embodiment 4

[Drawing 11] The block diagram for explaining operation of the image retrieval apparatus concerning a 5th embodiment of this invention

[Drawing 12] The flow chart for explaining a Prior art

[Drawing 13] The flow chart for explaining a Prior art

[Description of Notations]

- 1 Computer
- 2 Display
- 3 Video playback equipment
- 4 A/D converter
- 5 Control line
- 6 External storage
- 7 Input device
- 8a-8e Interface
- 9 CPU
- 10 Memory
- 11 Image input part
- 12 Face primary detecting element
- 13 Image output part
- 14 Scene change primary detecting element
- 15 Man-and-woman identification part
- 16 Age identification part
- 17 Expression identification part
- 18 Racial identification part
- 19 Characters list

- 20 Face database
- 21 Face image taking part
- 22 Feature extraction part
- 23 Characters identification part
- 24 Collating part
- 25 Retrieval object face image
- 26 The feature extraction part for video signals

---

[Translation done.]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

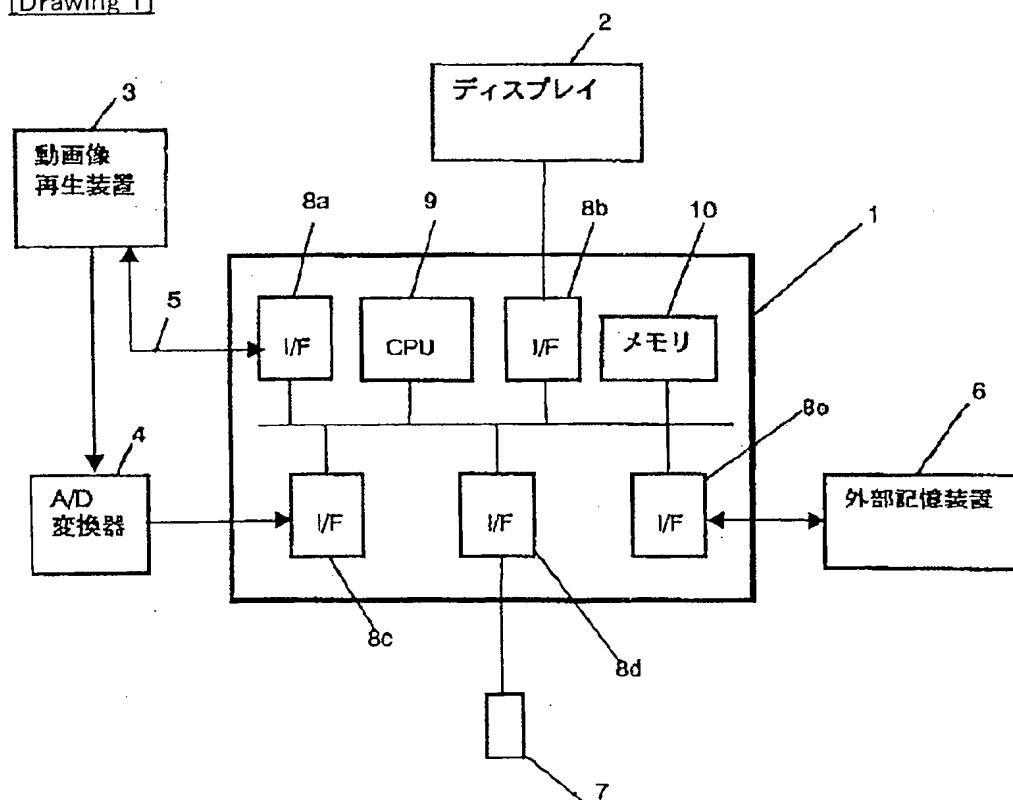
- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

DRAWINGS

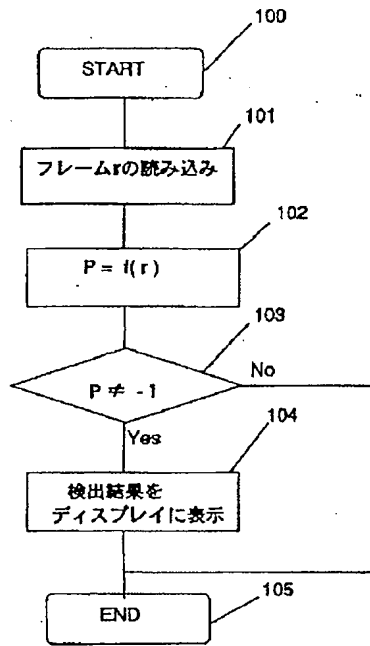
---

[Drawing 1]

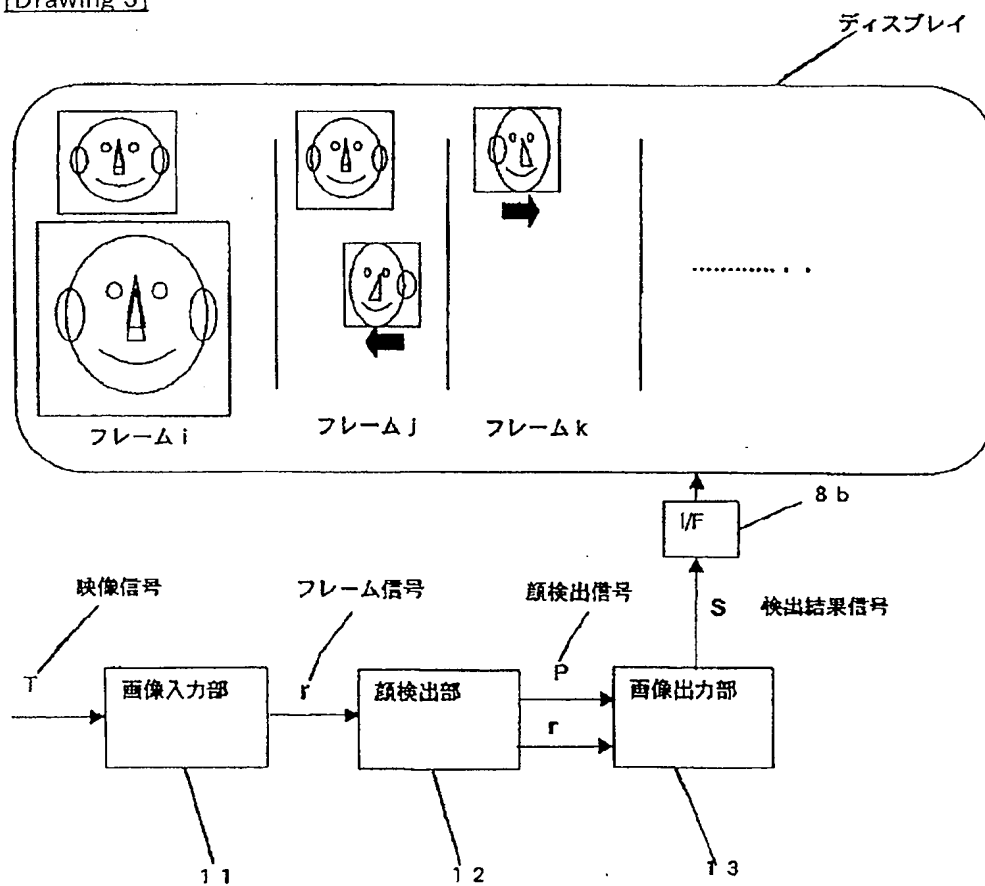


[Drawing 2]

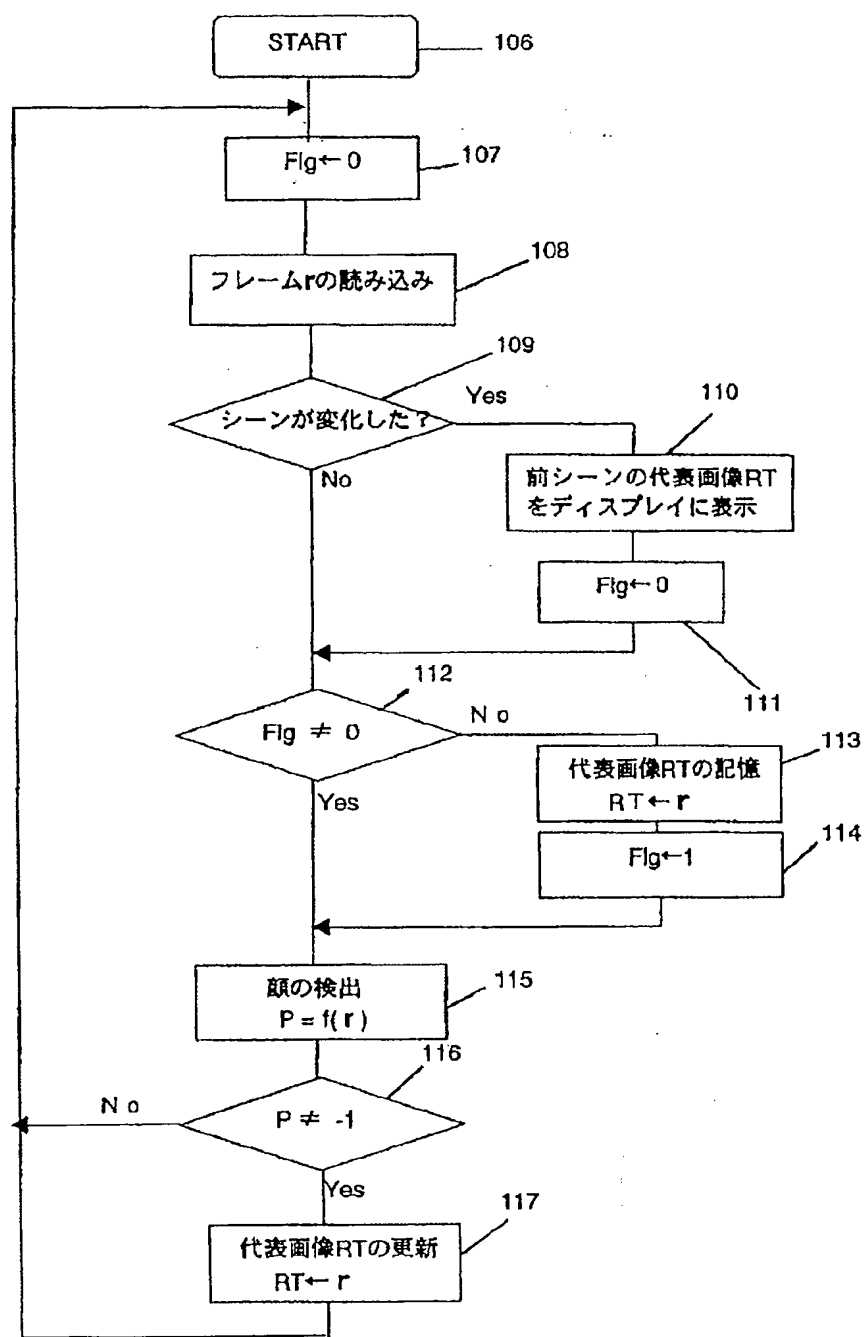




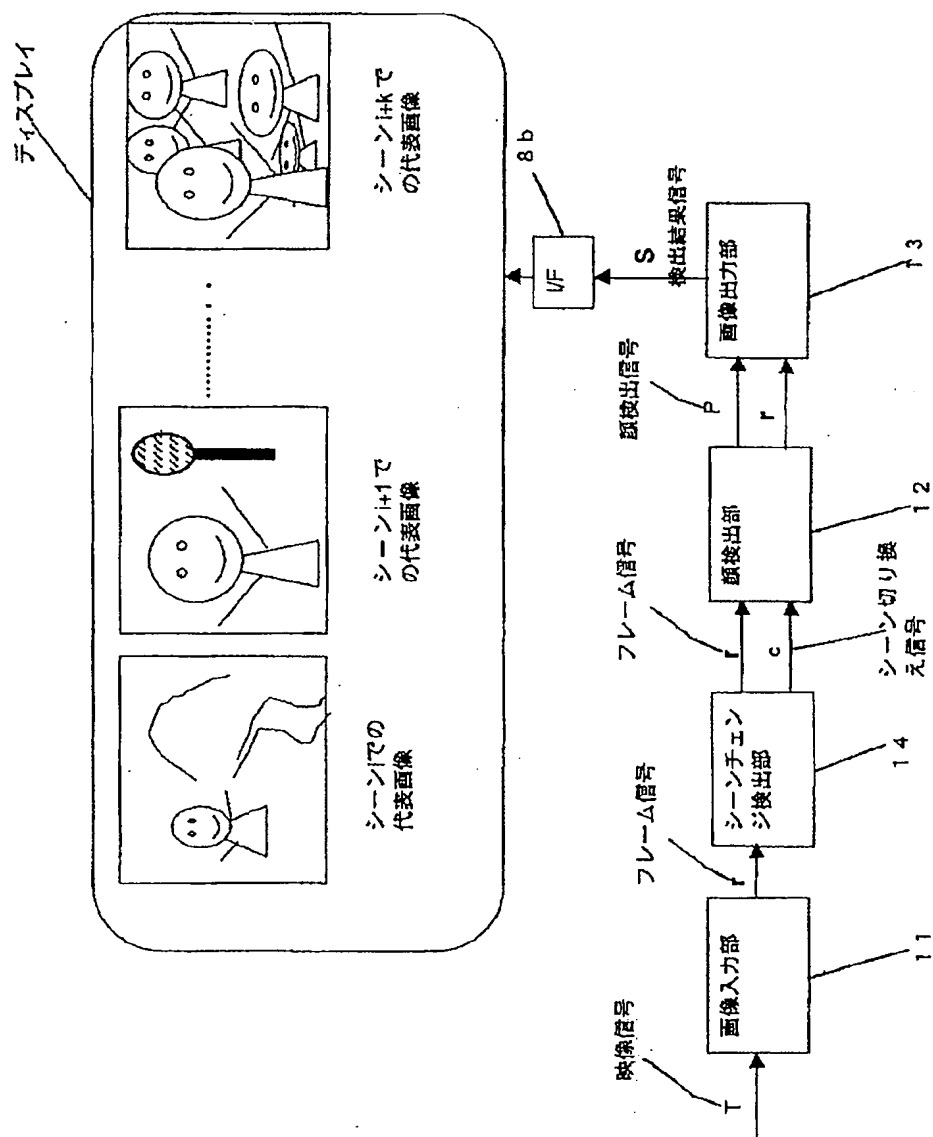
[Drawing 3]



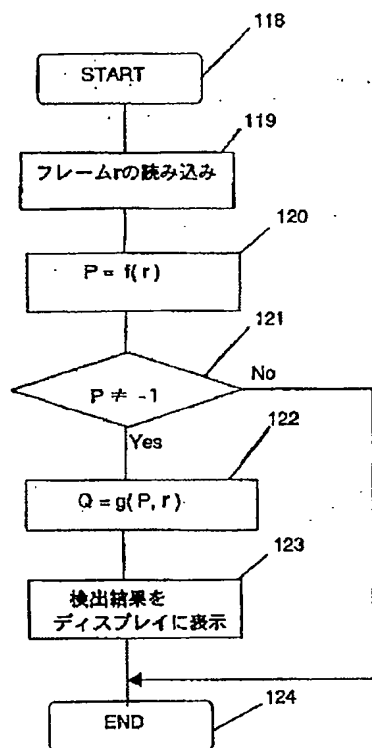
[Drawing 4]



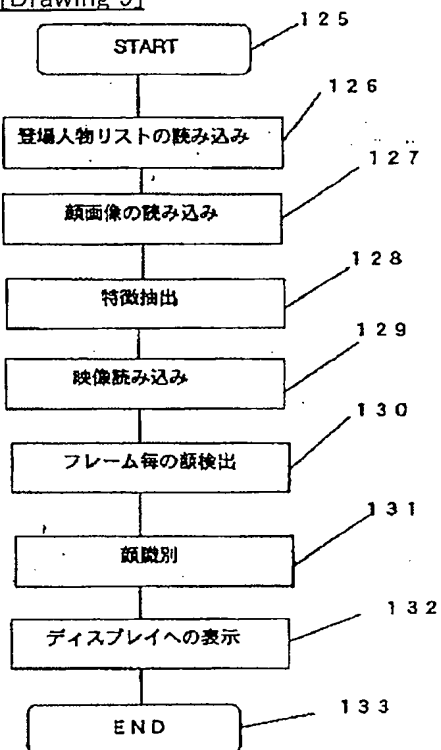
[Drawing 5]



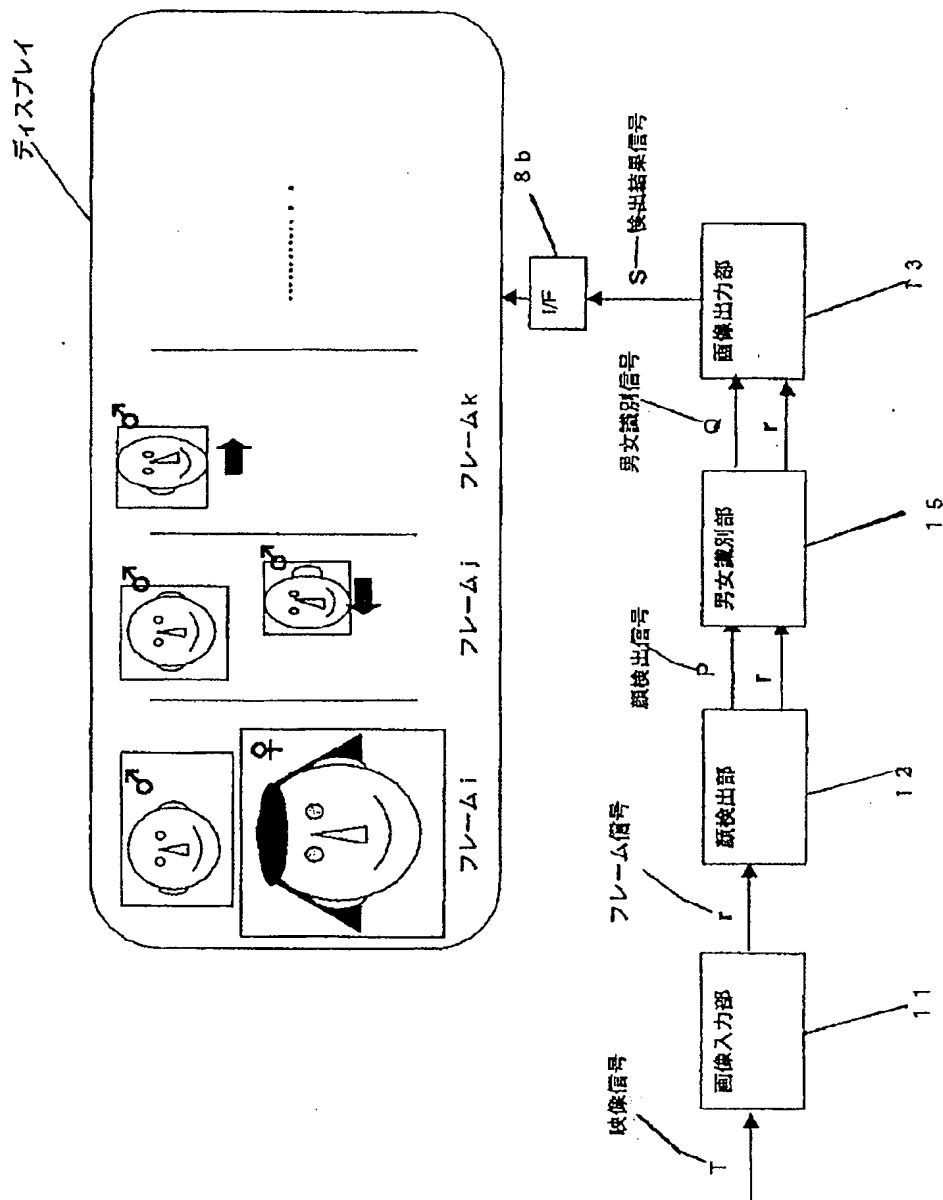
[Drawing 6]



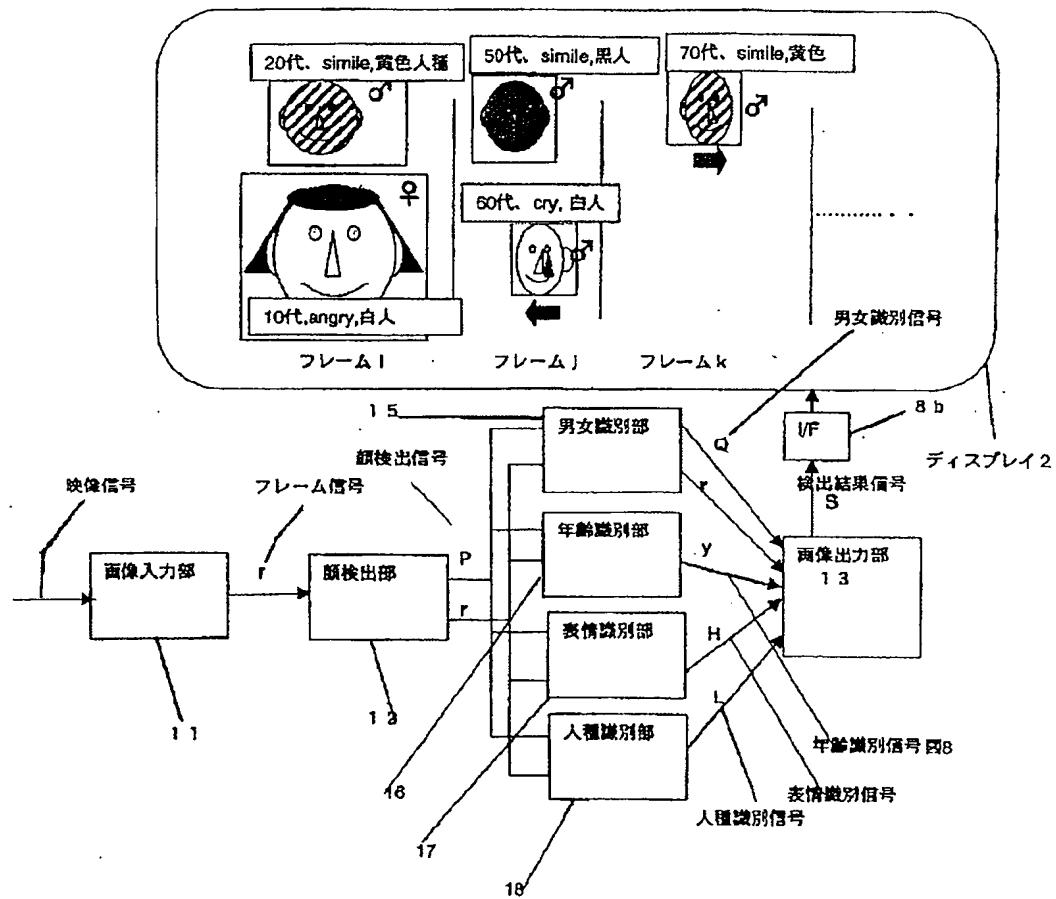
[Drawing 9]



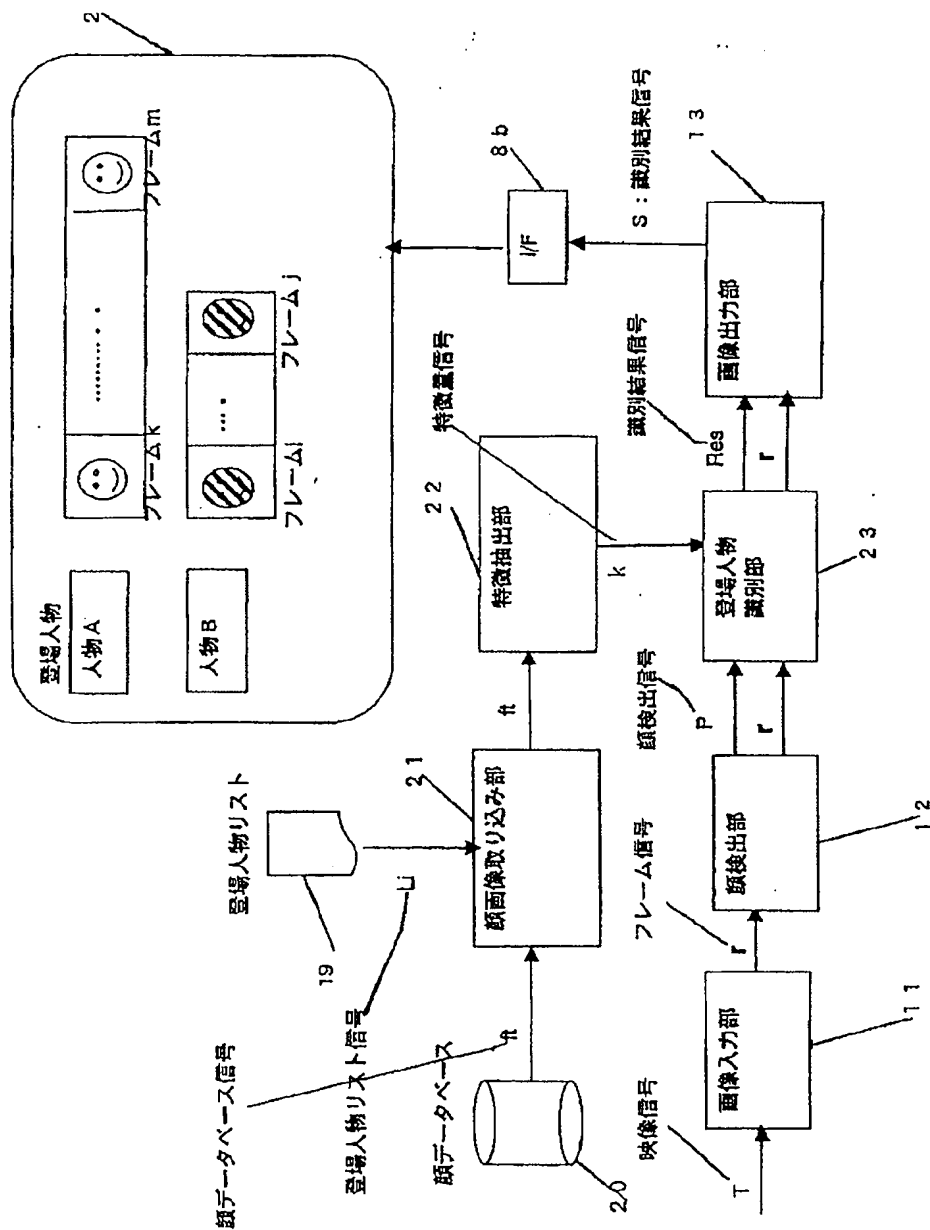
[Drawing 7]



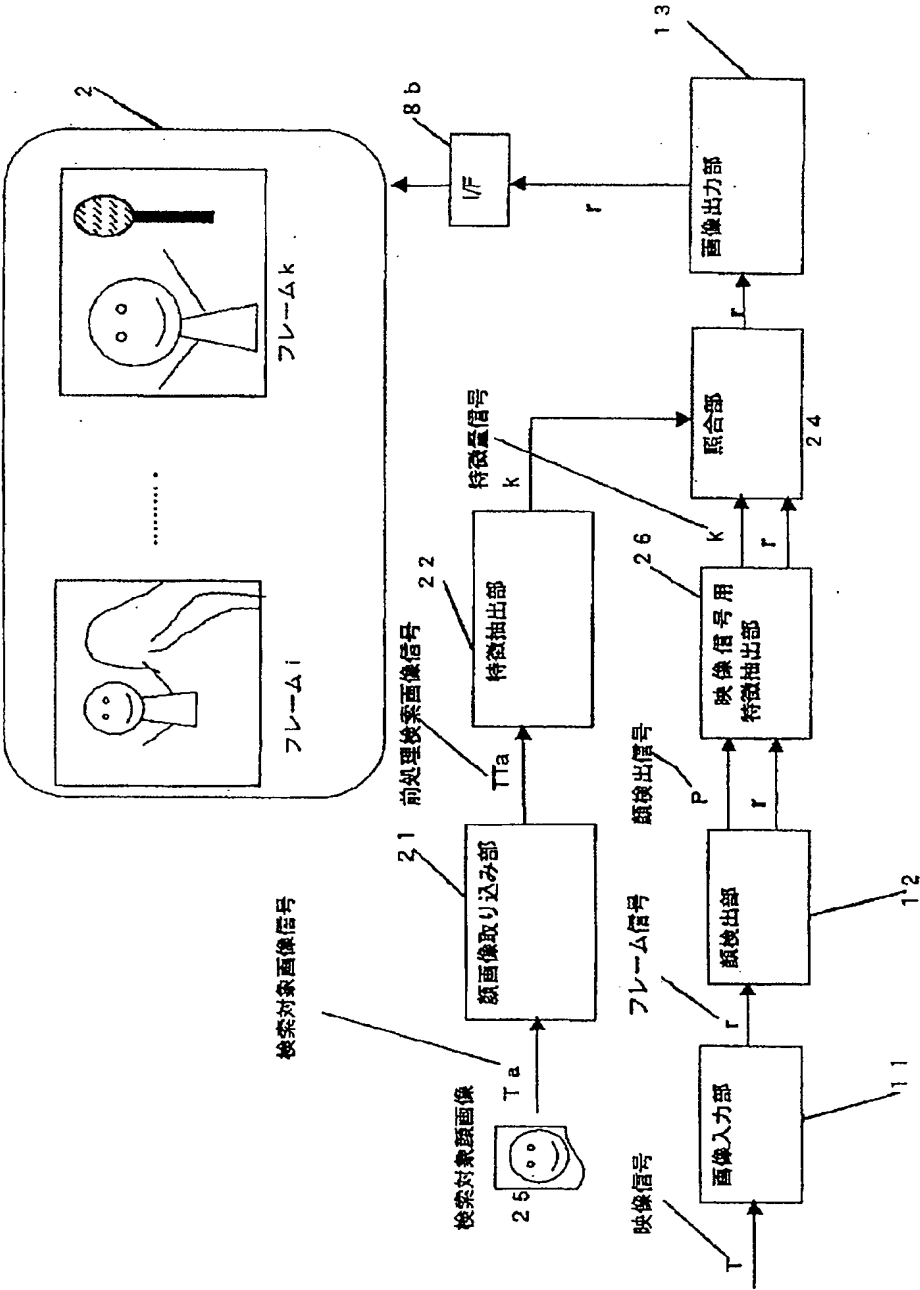
[Drawing 8]



[Drawing 10]

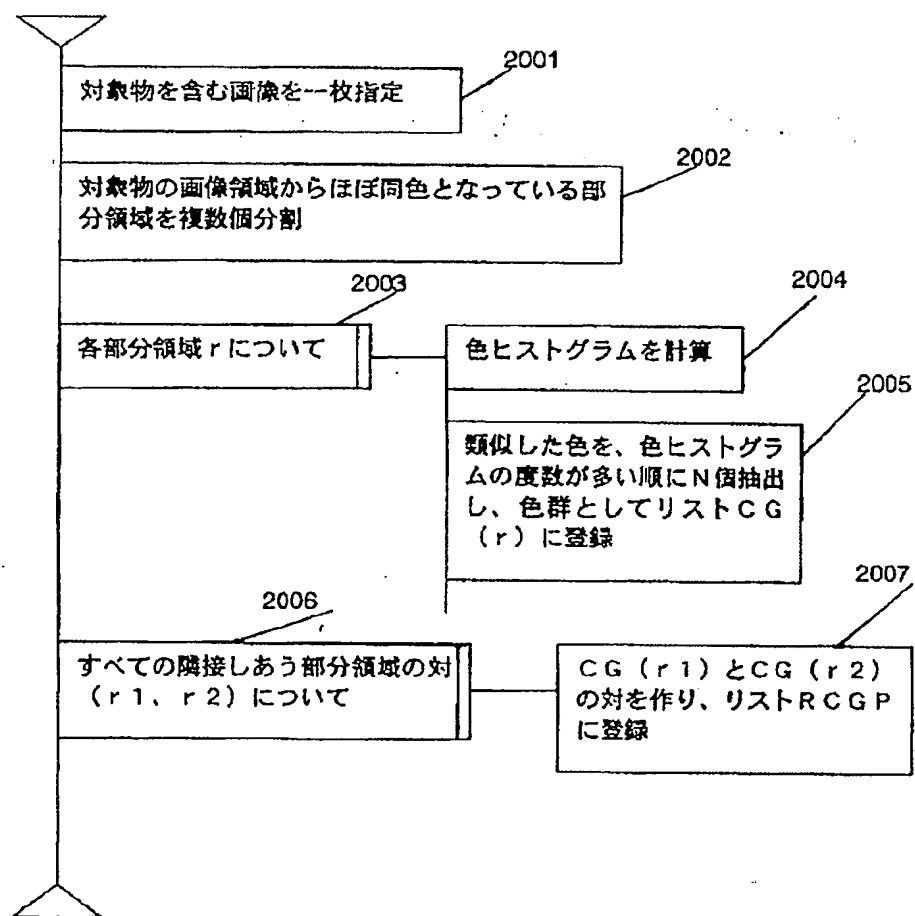


[Drawing 11]

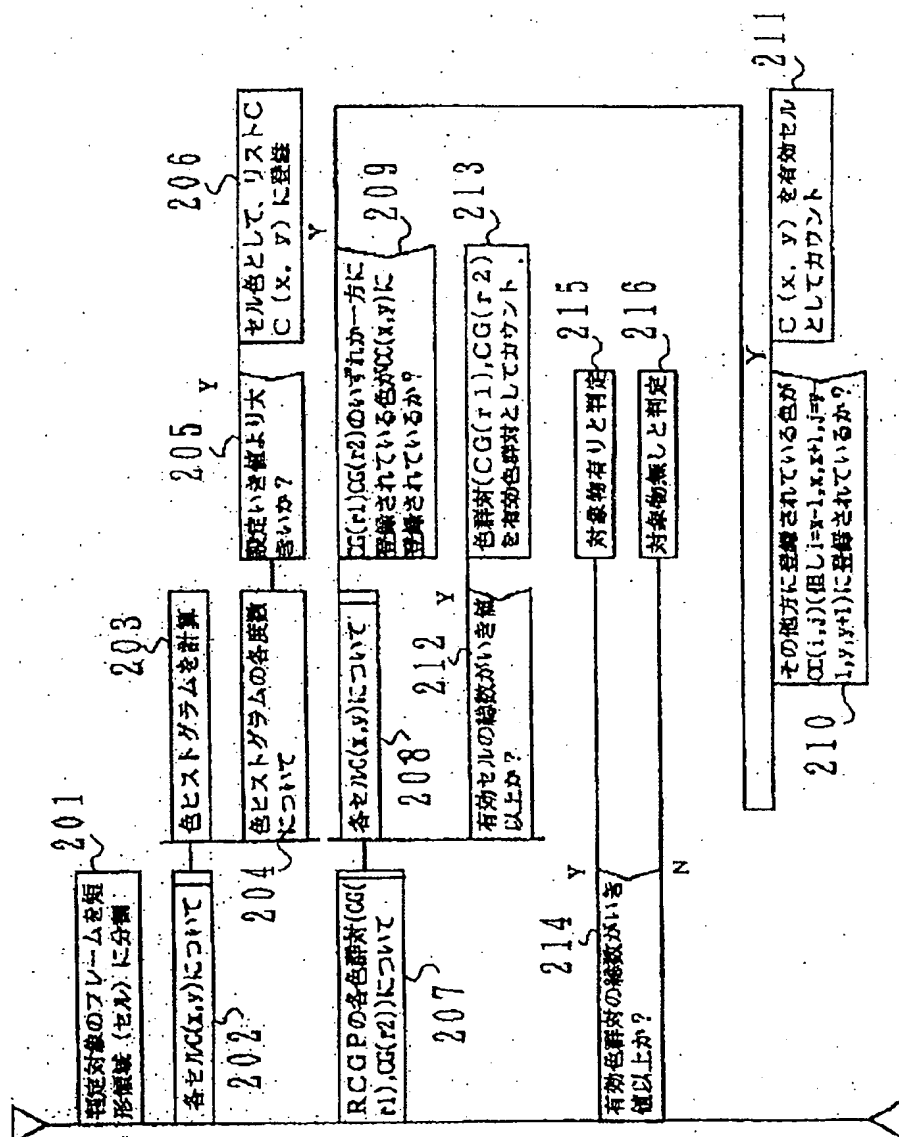


[Drawing 12]





[Drawing 13]



[Translation done.]